The potential value of early (1939-1967) upper-air data in atmospheric climate reanalysis

Hans Hersbach[1], Stefan Brönnimann[2], Leo Haimberger[3], Michael Mayer[3], Leonie Villiger[2], Joey Comeaux[4], Adrian Simmons[1], Dick Dee[1], Sylvie Jourdain[5], Carole Peubey[1], Paul Poli[6], Nick Rayner[7], Alexander Sterin[8], Alexander Stickler[2], Maria Antónia Valente[9] and Steven Worley[4].

1 Correspondence to: H. Hersbach, ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom. E-mail: hans.hersbach@ecmwf.int

2 Oeschger Centre and Institute of Geography, University of Bern, Bern, Switzerland

3 Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

4 National Center for Atmospheric Research, Boulder Colorado, USA

5 Direction de la Climatologie et des Service Climatiques, Toulouse, Météo-France

6 Centre de Météorologie Marine (CMM). Direction des Systèmes d'Observation (DSO), Météo-France, Brest, France

7 Met Office Hadley Centre, Exeter, United Kingdom

8 Russian Research Institute for Hydrometeorological Information - World Data Center (RIHMI-WDC), Obninsk, Russia

9 Instituto Dom Luiz, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

Abstract. In recent years a number of reanalysis datasets have been published that cover the past century or more, including the "Twentieth Century Reanalysis" 20CR and the European Reanalysis of the twentieth century ERA-20C. These datasets are widely used, showing the need for, and possible benefit of, reanalysis data products designed for climate applications. The 20th century reanalyses so far have assimilated only surface observations, and rely on independent estimates of monthly averaged sea-surface temperatures and sea ice concentrations as boundary conditions. While 20CR uses only observations of surface and sea-level pressure, ERA-20C additionally assimilates marine winds.

Here we describe an experimental reanalysis, referred to as ERA-PreSAT, which covers the period 1939-1967 and also assimilates historical upper-air data.

Assessments of this data set including comparisons with independent data show that (1) temperature biases in the northern hemisphere are largely reduced compared to reanalyses that assimilate surface data only, (2) concentration of 1940s upper air data in the northern extratropics created a strong interhemispheric asymmetry which is likely not realistic, (3) the forecast skill in the northern hemisphere has increased substantially compared to reanalyses that assimilate surface data only, (4) day-to-day and (in the northern extratropics) month-to-month correlations with independent observations (of total column ozone, upper-air data) increase over time, (5) interannual variability is well captured in the reanalysis, (6) a signature of the stratospheric Quasi-Biennial Oscillation is present as far back as the 1940s and (7) tropical cyclones are not well represented.

The encouraging results from the experimental ERA-PreSAT reanalysis underline that early upper-air data greatly contribute to our knowledge on the troposphere and lower stratosphere over the 20th century.

1. Introduction

Reanalyses are increasingly used for studying historic weather events and to assess multi-decadal variability in weather and climate. In recent years a number of reanalysis data sets have been published that cover the past century or more.

Examples are the “Twentieth Century Reanalysis” 20CR and 20CRv2c from the National Oceanic and Atmospheric Administration Cooperative and the Institute for Research in Environmental Sciences (NOAA/CIRES-CDC), Compo et al. 2011, and the European Reanalysis of the twentieth century ERA-20C of the European Centre for Medium-Range Weather Forecasts (ECMWF), Poli et al. 2016. These datasets are widely used, showing the need for, and possible benefit of, atmospheric datasets that cover the early part of the instrumental period prior to the satellite-dominated modern era. So far, the 20th-century reanalyses have only assimilated surface observations, in addition to relying on boundary conditions derived from monthly estimates of sea-surface temperatures and sea-ice concentrations. 20CR and 20CRv2c assimilate surface and sea-level pressure, whereas ERA-20C additionally assimilates marine wind. The European projects ERA-CLIM and ERA-CLIM2, which have produced the ERA-20C and more recently the coupled CERA-20C reanalysis, additionally have recovered a large number of historical upper-air data from analogue media and

prepared them for use in reanalysis (Stickler et al. 2014a, 2014b). Together with the previously published The Comprehensive Historical Upper Air Network (CHUAN, Stickler et al. 2010), millions of upper-air profiles are available for the first half of the 20th century, which up to now however have only been used for validation purposes (e.g., Compo et al. 2011, Brönnimann et al. 2012b, Stickler et al. 2015).

Here we present an experimental reanalysis, termed ERA-PreSAT, covering the years 1939-1967, which uses the same assimilation system as ERA-20C but additionally assimilates upper-air observations from the CHUAN historical dataset (Stickler et al. 2010) supplemented by data from the upper-air archives at the National Center for Atmospheric Research (NCAR). We describe these datasets and how they were ingested into the assimilation system. Then we evaluate some aspects of the atmospheric energy cycle. We evaluate the reanalyses using additional independent historical upper air data from the ERA-CLIM dataset (Stickler et al. 2014) that had become available later. In addition, we also use historical observations of total column ozone to evaluate the ozone estimates in the reanalysis datasets (Brönnimann and Compo, 2012).

The paper is organised as follows. The ERA-PreSAT reanalysis system and the ingested historical upper-air data are described in Section 2. The specifications of the datasets for comparison are presented in Section 3, which include one reanalysis product and three different upper-air reconstructions. In Section 4 various aspects of ERA-PreSAT are validated and/or compared with these datasets. The paper ends with a discussion and conclusion (Section 5).

2. The ERA-PreSAT forecast and analysis system

The ERA-PreSAT model and data assimilation system is equal to that of ERA-20C (Poli et al. 2016). It is based on version 38r1 (ECMWF 2013) of the Integrated Forecasting System (IFS), but at a reduced resolution compared to the configuration used for operational weather forecasting at ECMWF. Horizontal spectral resolution is T159 (around 125km globally) and there are 91 levels in the vertical from the surface up to 1 Pa (around 80km) with roughly 51 in the troposphere, 31 in the stratosphere and 9 in the mesosphere. The atmosphere is two-way coupled with an ocean-wave model and a land-surface model. A detailed description of the IFS can be found at https://software.ecmwf.int/wiki/display/IFS/CY38R1+Official+IFS+Documentation.

Boundary conditions at the sea surface are obtained from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST) version 2.1.0.0 (Titchner and Rayner 2014; Kennedy et al. 2015), and radiative forcing is mostly obtained from CMIP5 recommended datasets. It has been shown (ERA-20CM, Hersbach et al. 2015) that a model-only integration of these century-varying boundary conditions and forcing adequately represent the main low-frequency variability of the 20th century atmosphere (global warming, El Nino and La Nina events, the effect of major volcanic eruptions), without the need for the assimilation of synoptic observations.

The analysis uses four-dimensional data assimilation (4D-Var), with two inner loops at T95 horizontal resolution (around 210 km). Like ERA-20C, information on the first-guess background errors is obtained from a previously produced ensemble of 20th-century data assimilations (Poli et al. 2015). Details can be found in Poli et al. 2016.

2.1 The ERA-PreSAT observational data input

Like ERA-20C, ERA-PreSAT assimilates surface pressure and sea-level pressure data from the International Surface Pressure Databank (ISPD, Compo et al. 2010) version 3.2.6 and the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Woodruff et al. 2011) version 2.5.1, as well as marine wind reports from ICOADS. For details and information on the evolution over the 20th century of this observing system see Poli et al. (2015, 2016).

In addition, ERA-PreSAT assimilates historical upper-air data from several sources: the NCAR Upper Air Data Base Version 2 (UADB-2), the Comprehensive Historical Upper-Air Network Version 1.7 (CHUANv1.7), as well as observations digitized within the ERA-CLIM project (ERA-CLIM Version 0.9).

The UADB-2 data record contains the original records from a large number of sources, has been uniformly formatted, metadata are standardized and measurement units are consistent. A detailed description can be found at http://rda/ucar.edu/datasets/ds370.1/UADB-Doc.pdf. ERA-PreSAT uses data from 44 sources, as displayed in Table 1 of the UADB document, with the exception of sources 23 (China), 51 (NCDC 5420) and 151 (Russian Ships) which were not yet available in the UADB-2 format at the time of the preparations for ERA-PreSAT.

The CHUAN v1.7 data record (Stickler et.al. 2010) comprises 3,987 station records worldwide totalling about 16.4 million vertical profiles. It was also obtained from NCAR as two separate sources: a raw (‘r’, source 31) and a corrected dataset (‘c’, source 30). Both sets originate from the same data, though the latter contains some corrections, such as RAOBCORE v1.3 (Haimberger, 2007) temperature bias adjustments.

Note that CHUAN v1.7 also contains a large amount of data as monthly means for which the underlying individual observation profiles are not available (Stickler et al. 2010), most notably over the USA in 1939-1944. These data may however be used for independent validation.

The ERA-CLIM project included a large component of data rescue, inventoring, locating, imaging, and digitizing observation records on paper or microfilm. The ERA-PreSAT reanalysis uses a preliminary version of this data record (V0.9) and misses several collections that have been added since.

2.2 The assimilation of upper-air data in ERA-PreSAT

At ECMWF, all upper-air data as described above were converted to the format used at ECMWF for assimilating observations. Data were then presented to the assimilation system as either wind profiles (so-called PILOT) or multivariate profiles (so-called TEMP data: temperature and wind). Upper air humidity observations were not assimilatedsince they were suspected to have large biases at this time, even at heights below 300 hPa, as suggested by Dai et al. (2011). All observations were submitted to the assimilation and there was no attempt to remove duplicates beforehand (even though the CHUAN ‘c’ and ‘r’ data records contain duplicates). The removal of duplicates was left to the screening part of the assimilation system, where it could be decided which observations fit the background the best. This makes uses of the prior assimilation, and enables to discard duplicate observations. Besides duplicate removal, rejections are also made as a result of quality control, such as to screen data that depart too much from the background model, or data that cannot be fitted well by the analysis given all other constraints (e.g., an observation contradicting the others). All upper-air data are assimilated using pressure as the vertical coordinate. In case this

information is missing, pressure is estimated using the height-pressure relationship of the background, from the observation height, if available.

The observation errors represent the weight given to them in the assimilation, in balance with the background errors. The upper-air temperature and wind observation errors assumed in ERA-PreSAT follow the modern-day specifications. The assumed wind observation error is 1.6 m/s from the surface up to 850 hPa, then increases linearly with decreasing pressure to 2.3 m/s at 300 hPa, before decreasing to 1.8 m/s at 50 hPa, and then rapidly increasing to 2.7 m/s higher up. For temperature, the assumed observation error increases from 0.9 K at the surface to 0.6 K at 400 hPa, then inflates to about 1.3 K at 50 hPa and then quickly further increases up to 1.6 K higher up. After the production of ERA-PreSAT it was realized that these profiles had inadvertently been swapped in some occasions. In the IFS data assimilation system, PILOT data is expected as a function of height whereas TEMP data is expected as a function of pressure. However, for the historical data, all combinations occur. Unfortunately, PILOT data as a function of pressure was erroneously assigned observation errors as a function of height, while a similar error was made for TEMP data as a function of height.

Given the fact that the modern network is of (much) higher quality than the historical data (see Wartenburger et al. 2013), the historical upper-air data are probably assigned errors that are too small, and consequently given too much weight. Also, no bias corrections were applied to the observations, except for the CHUAN ‘c’ records that contain RAOBCORE v1.3 (Haimberger, 2007) bias adjustments. This absence of bias correction is suboptimal, since large systematic errors are known to exist in the observations. Known issues include for example a radiative warm bias for high temperature profiles, particularly over the Former Soviet Union (Grant et al. 2009) and systematic errors in wind direction for part of the early US pilot network (Ramella-Pralungo and Haimberger 2014).

Clearly, the absence of dedicated background errors, the mixing of all observational sources without any prior duplicate removal, the prescription of probably too small observation errors (plus occasional mix-up in the vertical), and the neglect of observation biases are duely acknowledged as non-optimal. However, these issues are difficult to spot and resolve upon first trial. Owing to project time constraints it was not feasible to correct these in a rerun of ERA-PreSAT within the limited time frame of the ERA-CLIM project. Nevertheless, given the ERA-20C baseline without any

upper-air assimilation, the rest of this paper demonstrates that the addition of upper-air data in ERA-PreSAT, albeit in a suboptimal framework, allows to assess the enhancement of the reanalysis product, far outweighing the suboptimal usage of these observations.

Figure 1 presents a time-line of the availability during the ERA-PreSAT period for the four upper-air records and the assimilated. Although the largest dataset is represented by UADB-2, it is seen that the recently digitized ERA-CLIM data improve considerably the availability of temperature soundings before 1943. Regarding data usage, there is quite some competition between the CHUAN-R and CHUAN-C sets for the obvious reason that these sets are based on the same observations. It appears that the screening in general favours the CHUAN-C dataset, which indicates that corrected observations may generally be more consistent with the background than uncorrected observations. Also, it is very reassuring that most of the ERA-CLIM data is actually selected for assimilation, as it indicates most likely independent, new, timeseries, and not inconsistent with other observations (either or both such condition would have resulted in many data rejected).

The evolution of the global coverage is illustrated in Figure 2, which shows the average daily number of actively used observations accumulated in 5x5 degree boxes for the years 1943, 1950, 1957 and 1964. From this it emerges that initially temperature records were very sparse and mainly concentrated over Northern Europe and Russia (from 1939), South Korea and China (from 1942), Northern South America and Northern Australia (from 1943). Data from India and Pakistan is available between 1939 and 1941 (not shown). From 1946 onwards radiosonde data over the United States of America (US) became available, the Russian network expanded, while the West-European network expanded rapidly in 1948. From around 1957 more soundings became available from South America and South Africa. The availability of wind profiles was already quite good over the US from 1939, Southern Europe, India, Korea and East China. More sparse data was available from South America and Central Africa. In 1943 there was a boost in wind soundings over the Eastern part of the US, likely related to the 2nd World War. There was a further boost over the entire US from 1948. From 1962 there has been a rapid increase in the availability of both upper-air wind and temperature observations over the northern hemisphere oceans.

3. Data used for comparison

3.1. 20CR and other Reanalysis data

ERA-20C and ERA-PreSAT are compared with Twentieth Century Reanalysis (20CR) version 2. 20CR which is a 3-dimensional, 6-hourly global atmospheric dataset and is based on the assimilation of surface and sea level pressure observations into the US National Centers for Environmental Prediction Global Forecast System atmosphere/land model (NCEP/GFS, Saha et al. 2010). It is run at a resolution of T62 in the horizontal and 28 hybrid sigma-pressure levels in the vertical (Compo et al. 2011). Monthly mean sea surface temperature and sea ice concentration from the HadISST dataset (Rayner et al. 2003) are used as boundary conditions. The assimilation was performed using a variant of the Ensemble Kalman Filter, and 56 ensemble members were used. In this study, however, we only address the ensemble mean.

For further comparisons we also used the reanalyses ERA-40 (Uppala et al. 2005), ERA-Interim (Dee et al. 2011) and JRA-55 (Kobayashi et al. 2015).

In addition, several comparisons are made to the ERA-20CM (Hersbach et al. 2015) model integration.

3.2. Statistical reconstructions

In addition to reanalysis data, we also compare ERA-PreSAT with monthly statistical reconstructions of global upper level fields. Three different reconstructions are used, termed BL (Brönnimann and Luterbacher 2004), REC1 (Griesser et al. 2010) and REC2 (Brönnimann et al. 2012a) hereafter.

All three reconstruction approaches are based on principal component (PC) regression. All of them use historical upper-air and surface data (sea-level pressure and station temperatures) as predictors and calibrate these data against reanalysis fields for the past few decades. BL focused on the 1939-1945 period and produced fields of temperature and geopotential height (GPH) at six levels (850, 700, 500, 300, 200, and 100 hPa) for the northern extra-tropics by calibrating against NCEP/NCAR reanalysis (Kistler et al. 2001). REC1 uses the same method and same output fields (but now global, *i.e.*, reconstructions were produced separately for the regions 15°-90° N, 20° S-20°N, and 90° S-15° S, and back to 1881), but large amounts of additional

upper air data. Moreover, it was calibrated against ERA-40. REC2 has the same output levels but now also zonal and meridional wind components. It also uses PC regression calibrated against ERA-40. However, REC2 is a grid-column-by-grid-column reconstruction. For each grid column, only predictors from a cone of influence around that grid column are considered. A minimum amount of upper-level observations is required, and weights are attributed such that upper-air data contribute at least 50%. Thus, unlike in BL or REC1, no stationarity of large-scale spatial patterns is assumed. However, grid columns away from upper-air observations have no data, and the resulting fields are not necessarily smooth or physically consistent. REC2 is thus more akin to an interpolation of upper-level observation data.

3.3. Total column ozone data

Historical total column ozone data provide an interesting opportunity to independently evaluate the performance of the historical reanalysis datasets. We use historical ozone data from the World Ozone and Ultraviolet Radiation Data Center (WOUDC) from 1939 to 1963. Among the series are long series such as the well-studied series from Arosa, Switzerland (46.8° N, 9.7° E; Staehelin et al. 1998) or Dobson's original series from Oxford, UK (51.8° N, 1.2° W; Vogler et al. 2007), but also many shorter series (see Brönnimann et al. 2003, for an overview). We use the same selection as in Brönnimann and Compo (2012), but some of the stations used in that study do not have data after 1939. Table 2 gives a list of the stations and number of daily values.

4. Results

4.1 Forecast skill.

As described above, the ERA-PreSAT reanalysis was set-up in exactly the same way as ERA-20C. The only difference is the usage of upper-air data. This, therefore, enables a very clean assessment of the impact of these data. For both ERA-20C and ERA-PreSAT, a ten-day forecast had been integrated from each 00UTC analysis. This allows for the assessment of the potential value of the upper-air data on forecast skill in case they had been available in near-real time and the current data assimilation scheme had been available at the time. Resulting forecast scores for the anomaly correlation coefficient of the geopotential at 500 hPa height are presented in Figure 3.

It displays the average number of days for which the forecast had been excellent (90%), good (80%) and on the edge of just being better than climatology (60%) over the northern hemisphere (top) and Europe (lower panel). It shows an initial small or neutral impact for the early 1940s but already in the 1950s there is a dramatic gain of about 1.5 days. Especially the sudden large increase in forecast skill around 1948 over Europe is noteworthy. It coincides with the large increase of wind soundings over the US. A similar, though smaller pattern is seen for 1944. A synoptic example of this improvement of forecast skill has been studied for the D-Day landing in June 1944 (Simmons et al., 2015). The skill of ERA-PreSAT outperforms the ERA-40 reanalysis (Uppala et. al. 2005, dashed curves), which latter is better than ERA-20C. This comparison should be handled with some care, since due to the difference in the length of assimilation windows (as detailed in the caption of Figure 3), the ERA-40 forecasts could have been disseminated 18 hours before those for ERA-20C and ERA-PreSAT.

The initial decline in scores over the northern hemisphere is thought to be the result of the decrease in the availability of observations during the 2nd world war. It especially applied to North America (not shown). Apparently the first upper-air soundings were not able to reverse this picture. Over Europe, however, the positive impact emerged right from the early 1940s, which, again, is likely the result of the availability of (on-average upstream) wind profiles over the US. It is also interesting to note that for Europe the best forecast scores are achieved around 1960 and that afterwards some form of decline is apparent. Such decline is not visible over the North America and East Asia (not shown).

4.2. Basic energetic considerations

To explore the evolution of the energetic state of the atmosphere in the different ECMWF reanalyses, we consider vertically integrated total energy from the different datasets averaged over different periods of time, choosing 2000-2009 as reference period. We choose 1939-1944 and 1961-1966 representing the early and late periods of ERA-PreSAT, respectively. In order to remove the impact of different model topographies and differences in the mean circulation, i.e. regional mean surface pressure values, we present a comparison of zonal mean vertically averaged total energy from the respective datasets in Figure. 4a.

The early ERA-Interim period (1989-1999) shows only slightly lower values compared to the reference period (2000-2009), except for the Arctic, where strong warming was present during the 2000s. Atmospheric energy in ERA-PreSAT exhibits remarkable differences compared to ERA-Interim. During the early period (1939-1944), ERA-PreSAT shows a reasonable energetic state north of about 30N where upper-air observations were available already at that time. The increasing number of upper air observations all over the globe is reflected in a higher energetic state of ERA-PreSAT at all latitudes by 1961-1966. During that time, this reanalysis is already in very good agreement with the early ERA-Interim period (1989-1999). This indicates that the climatological state of the assimilating model without upper-air observations is energetically too low.

This impression is confirmed when examining ERA-20C and ERA-20CM which do not assimilate any upper-air observations. Although each of these datasets shows a relative increase of specific energy over time, consistent with global warming, both reanalyses exhibit values about 0.5% low when compared to ERA-Interim during the 2000-2009 period. As the by far largest amount of atmospheric energy is represented by enthalpy, we conclude that the assimilating model of ERA-20CM, ERA-20C, and ERA-PreSAT has a cold bias, translating to about 1K for 2000-2009.

The gradual evolution of the atmospheric state in ERA-PreSAT can also be seen from the zonal mean temperature structure (Figure 5). While during 1939-1944 tropospheric temperatures in ERA-PreSAT are higher than in ERA-20C mainly north of 30N (Figure 5a), positive differences are present globally during 1961-1966 (Figure 5b). As a reference panel c) shows ERA-Interim-ERA-20C differences in the period 1980-1989. The pattern looks quite similar to panel b) and shows the cold bias in ERA-20C. It indicates that the state of ERA-PreSAT in the 1960s was already quite realistic, except perhaps in the Southern Ocean and Antarctica when there were still too few observations. Figure 5 also shows that differences are comparatively small in the lower troposphere, indicating that differences in SSTs can be ruled out as a cause for the found differences (see also Figure 4b). Moreover, Figures 5b and 5c imply that the modelled troposphere is statically too unstable when no upper-air observations are assimilated. The hemispheric asymmetry of observations going into ERA-PreSAT in the early period obviously affects analysed stratospheric temperatures, with too cold (warm) temperatures above about 300hPa in the Southern (Northern) Hemisphere when compared to ERA-20C (Figure 5a). The stratospheric temperature differences

also explain the very low specific total energy of ERA-PreSAT south of about 45S during 1939-1944 (see Figure 4a). Since the differences in stratospheric temperatures are not present for the later period (Figure 5b), we speculate that the hemispherically asymmetric atmospheric state in ERA-PreSAT during the early period adversely affects stratospheric circulation, which leads to the described spurious temperature patterns.

Besides the temporal evolution of the number of upper-air observations, their spatial distribution is important too. An uneven distribution of observations can set up unrealistic gradients of atmospheric energy, which likely affects atmospheric energy transports. Cross-equatorial energy transports as presented in Figure 6 are a quite sensitive diagnostic in this respect. ERA-20CM exhibits quite stable negative values around -0.17 PW, in good agreement with ERA-Interim (~ -0.25 PW; compare also Mayer and Haimberger 2012). Transports from ERA-20C in the 1940s are too weak but approach ERA-20CM values later in the period. Transports from ERA-PreSAT have the wrong sign in the 1940s, but approach ERA-Interim values in the 1960s. This behavior is likely caused by the unrealistic interhemispheric temperature gradients in the 1940s seen in Figure 5 which gradually disappear in the 1960s. The reason for the changing temperature gradients is mostly the changing coverage of upper air data. Surface wind and pressure observations probably also contribute to the unstable cross-equatorial energy transport in ERA-PreSAT, as ERA-20C shows similar variations.

4.3. Comparison with independent observations and reconstructions

4.3.1. Comparison with ERA-CLIM upper-air data

Apart from the NCEP/NCAR 50-Year Reanalysis (Kistler et al. 2001) and a few very short term reanalyses, ERA-PreSAT is the only reanalysis dataset where upper air data before 1958 have been assimilated. Despite some technical issues noted above, it turned out that the early data are generally of high quality and have a profound influence on the atmospheric state.

Figure 7 shows that departures between radiosonde observations and ERA-PreSAT 12 hour forecasts (referred to as obs-bg) for a given site over the US are much smaller (0.52K standard deviation at the 200hPa level) than departures between observations and ERA-20C analyses (obs-an 1.21K at this level).Note that for ERA-preSAT obs-bg

departures have been chosen since the background forecasts at a given radiosonde site are largely independent of the respective radiosonde observations (which is not true for the ERA-preSAT obs-an since the radiosonde data have been assimilated). For ERA-20C the analysis data are completely independent of radiosonde observations and therefore obs-an departures can be used. It is also noteworthy that the mean of the departures is much smaller for ERA-preSAT, whereas ERA-20C seems to have a cold bias compared to the radiosonde observations.

The small obs-bg departures highlight the overall good quality of the early radiosonde data and also indicate a good short term predictive skill of the ERA-PreSAT assimilating model. The small departures enable much more efficient detection of potential breakpoints in the radiosonde observation records than the ERA20C departures. In this particular case the change from BENDIX-FRIES to VIZ radiosondes 1957 has caused a shift of one degree that is barely detectable from the from ERA-20C obs-an departures using a popular homogeneity (SNHT, Alexandersson, 1986; Haimberger 2007) but is very well detectable from ERA-PreSAT obs-bg departures. The shift in 1948 also coincides with metadata for this station according to the CARDS data set (Eskridge et al. 1995).

Since the difference series between ERA-PreSAT 12h forecasts and ERA-20C (not shown) indicates no breaks in 1948, 1950 and 1957 the detected breaks are likely caused by changes in the radiosonde records. This and many further examples indicate the high potential of these reanalysis departure time series for automatic data quality control and homogenization of early aerological observations

4.3.2. Comparison with daily total column ozone anomalies

In order to address the reliability of day-to-day variability in ERA-PreSAT and other reanalyses (ERA-20C, 20CRv2), we compared total column ozone to observations. For this purpose we subsampled the reanalyses to the observations (both in time and space) and removed for each station the annual cycle by least-squares fitting and subtracting the first two harmonics of the day of year. Maps of the correlations are shown in Figure 8. The general structure shows highest correlations over the northern mid-latitudes (a feature known since the 1920s; Dobson and Harrison 1926), where column ozone and atmospheric dynamics are intrinsically linked (Vaughan and Price, 199, Orsolini et al. 1998; Barriopedro et al. 2010). Correlations drop towards the

Arctic and particularly rapidly towards the monsoon regions and the tropics, where day-to-day variability of total column ozone is smaller and less strongly linked to the circulation near the tropopause and hence is more difficult to capture in a reanalysis (Compo et al. 2011). This behaviour is known from previous studies.

The anomaly correlations are generally high (note that also the historical observations from 1939-1963 are far from perfect). Our previous work has shown surprisingly high correlations between 20CR and historical total column ozone data (Brönnimann and Compo, 2012). Here we find even higher correlations with the new ERA reanalysis datasets. ERA-20C shows slightly (but consistently) higher correlations than 20CRv2. Interestingly, correlations are generally highest for ERA-PreSAT, where values up to 0.8 are found over northern Europe. The improvement over the Arctic and over the Indian monsoon region also is particularly noteworthy. This analysis shows that assimilating upper-air data considerably improves the atmospheric fields at higher levels.

4.3.3. Comparison of month-to-month variability in upper-air data

In a next step we analysed month-to-month variability. For this analysis we used monthly upper-air data from the USA, where a relatively dense radiosonde network was operating since 1939, but most series are only available as monthly means until 1945 and hence were not assimilated (Brönnimann 2003). We again subsampled the reanalyses as well as the reconstructions to the station locations and subtracted the mean annual cycle based on the 1939-1944 period. The number of available months was between 48 and 66, which is sufficient for statistical analyses. Table 3 shows the anomaly correlations for four selected stations covering tropical latitudes to midlatitudes); similar results were found for other locations. Note that these monthly data were used in all three reconstructions, hence they are not independent.

ERA-PreSAT shows generally higher correlations for geopotential height and for the 700 hPa level than the other renalayses. At midlatitudes (Sault St. Marie, 46.5° N, Washington DC, 38.9° N) it has the highest correlations of all reanalyses. In the tropics (Islas Santanilla, 17.5° N) all reanalyses become much worse except at the lowest level. Also one of the reconstructions (REC1) becomes worse while the other two (BL2004, REC2) still yield high anomaly correlations that even increase with altitude. Among the reanalyses, ERA-20C performs better than the others. Results for

459 the subtropical location (Miami, 25.8° N) are in between those from the midlatitudes
460 and for the tropics.

461

462 **4.3.4. Interannual variability**

463 For the analysis of interannual variability, we focused on Arctic temperature, which
464 showed peculiar changes between the 1930s (the early twentieth century Arctic
465 warming) and the 1960s (a cold period in the Arctic). We presented a systematic
466 assessment of reanalyses and reconstructions with respect to Arctic temperature
467 profiles in a previous paper (Brönnimann et al. 2012b) and here would like to report
468 on ERA-PreSAT in this respect (see also Wegmann et al. 2016).

469 The focus is on 700 hPa temperature in winter (Dec.-Feb.), which is the season and
470 level where we expect atmospheric circulation changes to have an impact on Arctic-
471 wide temperatures. Table 4 show correlations between different datasets (note that
472 here we have only one reconstruction, REC1, as REC2 does not completely cover the
473 Arctic and BL2004 ends in 1947, thus leaving too few degrees of freedom) for the
474 winters 1939/40 to 1966/67. For this comparison REC1 was extended from 1957 to
475 1967 using ERA-40.

476 Although none of the datasets stands out, Table 4 shows that ERA-PreSAT improves
477 over ERA-20C as it shows higher correlations with the other two datasets. ERA-
478 PreSAT and 20CRv2 are similar in that respect. Note that 20CRv2 has an error in the
479 specification of sea ice, which might affect the results. The error is fixed in the latest
480 version v2c, which is not yet published and therefore not systematically analysed here
481 (correlations are very slightly higher).

482 Anomalous climatic conditions prevailed in the northern extratropics in the late
483 winters of 1940-1942. Arguably initiated by El Niño conditions in the tropical Pacific,
484 the midlatitude upper troposphere and stratosphere exhibited a strong climatic
485 signature expressing a weakened polar vortex and warm lower stratosphere over
486 northern Siberia (Brönnimann et al. 2004). At the surface, cold winter dominated over
487 northeastern Europa (even affecting the Second World War) while winters were warm
488 in Alaska. Contrasted against the neighbouring winters of 1939, 1943 and 1944, the
489 average of the 1940-1942 winters is among the strongest signals in the climate system
490 on multiannual time scales. The anomaly is partly reproduced from sea-surface

491 temperatures alone. It is therefore an easy starting point for comparing the different
492 data products. We expect strong anomalies in the upper troposphere and stratosphere.

493 In fact, the difference between January-April averages of 1940-1942 minus 1939,
494 1943, and 1944 (Figure 9) reveals a characteristic structure in the upper troposphere
495 and lower stratosphere that appear in a very similar manner in all datasets.
496 Interestingly, except for the model simulations (ERA-20CM), all datasets show even a
497 stronger signal than REC2004, on which our original publication was based
498 (Brönnimann et al. 2004). All reanalysis datasets show a pronounced warming of the
499 polar stratosphere akin of sudden stratospheric warming events (SSWs). More
500 frequent SSWs were suspected from observations, which however are too scant and
501 the suspected dates of SSWs do not fit well with those in ERA-PreSAT (not shown).
502 ERA-PreSAT shows a somewhat stronger warming and polar vortex response than
503 ERA-20C and 20CRv2, respectively. Overall, all datasets pass this first test and at the
504 same time further corroborate the abnormality of climate during the period 1940-
505 1942.

506

507 **4.4. The Quasi Biennial Oscillation**

508 The representation of the Quasi Biennial Oscillation is a useful benchmark for
509 reanalysis datasets. In full reanalysis it is well represented but it poses a tough
510 challenge for surface data only reanalyses as well as climate models since it is
511 maintained by complicated wave interaction mechanisms (Baldwin et al. 2001).

512 ERA-PreSAT is almost a surface data only reanalysis in the early 1940s, particularly
513 in the Tropics, with the amount of upper air data gradually increasing also in the
514 tropics in the late 1940s and early 1950s. As such one can observe the transition of the
515 QBO state from a purely modelled one to a state well constrained by upper air data.

516 While there are many measures to quantify the QBO, we use here the zonally
517 averaged zonal wind at the 50 hPa level averaged between 20N and 20S as a proxy.
518 This quantity can be reliably estimated from reanalyses as well as from relatively
519 sparsely distributed radiosonde stations. Figure 10 indicates that ERA-PreSAT has a
520 reliable QBO pattern back to the early 1950s, which is a significant advance
521 compared to what was available before from reanalyses but still leaves room for
522 improvement since statistical reconstructions of the QBO are available back to the
523 early 1900s (Brönnimann et al. 2007). In the 1940s, while there is still some upper air

data, ERA-PreSAT has difficulty to reproduce the QBO amplitude as do the surface data only reanalyses. The inclusion of newly digitized data from Meteo-France as well as better representation of the QBO in future versions of the assimilating model may help to extend the period of realistic QBO representation in reanalyses back even further.

4.5. Tropical cyclones

One of the main applications of historical reanalyses is the analysis of past weather extremes, either in order to improve statistics, for analysing decadal variability in extremes, or in search of analogues for a present-day extreme. One example for the latter could be Typhoon "Cobra", which struck the United States Pacific Fleet about 480 kilometres east of the Philippine island Luzon on 18 December 1944, inmidst of World War II, killing about 790 people. This case is an interesting historical precedent for typhoon Haiyan in 2013 and therefore taken as an example. We compare the three reanalyses with a historical weather chart from NOAA (see also Feuchter et al. 2014).

In Figure 11 we show the fields for 18 Dec. 1944, 6 UTC. All reanalyses show at least a slight depression east of the Philippines. However, a tropical cyclone is only seen in 20CRv2, and also here the core pressure is much higher than indicated on the synoptic chart. As mentioned in Poli et al. (2016), the ECMWF variational quality control scheme of the observations has led to inadvertent exclusion or downweighting of many best-track reports or other tropical cyclone data (see also Poli et al. 2015).). This is also the case for “Cobra”, where the tropical cyclone reports are also excluded by the background check. This affects the representation of “Cobra” both in ERA-20C and ERA-PreSAT. Hence, although a low pressure value of 988 hPa is presented to all assimilation systems near the typhoon, values drop only to 1004.6 hPa and 1005.1 hPa in ERA-20C and ERA-PreSAT, recpectively, but to 999.6 hPa in 20CRv2. Even though during this period, upper-air data from neighbouring islands were assimilated into ERA-PreSAT, the assimilation is not improved near the storm. More detailed analyses of additional tropical cyclones will be presented in a forthcoming paper.

5. Conclusion

In this paper it is concluded that early upper-air data has great potential in improving our knowledge on the troposphere and lower stratosphere in future climate reanalysis. This was demonstrated by the production of a ERA-20C type reanalysis that covers

the period 1939-1967 and in addition to surface information had assimilated upper-air temperature and wind from three historical datasets. The analyses of this experimental reanalysis, ERA-PreSAT, and comparison with existing products and with independent upper-air observations, total column ozone, and other meteorological variables show the following:

- Biases in the northern hemisphere are largely reduced compared to surface data only reanalyses.
- The strong concentration of early upper air data to the northern extratropics in the 1940s created a strong interhemispheric asymmetry which is likely not realistic. This issue can presumably addressed by reducing the overall cold bias of the assimilating model and by correcting the generally warm biases of the assimilated early radiosonde data.
- The forecast skill in the northern hemisphere has increased substantially compared to surface data only reanalyses.
- Day-to-day and (in the northern extratropics) month-to-month correlation with independent data (total column ozone, upper-air data) increases.
- Interannual signals in the northern sub- and extratropics are well captured in all products.
- ERA-PreSAT display a signature of the stratospheric Quasi-Biennial Oscillation back to the 1940s.
- Like ERA-20C, tropical cyclones are not well represented in ERA-PreSAT.

ERA-PreSAT is an experimental product and will not be made available via a public data portal. Its prime purpose was to explore the usage and impact of early upper air data in climate reanalysis. Some short-cuts and errors had been made regarding the data ingestion. Examples are the non-optimal weight assigned to the observations, and the occasional mix up of such weights in the vertical. Despite such sub-optimal choices, the results of this experimental reanalysis are found to be very promising.

Overall ERA-PreSAT has shown that assimilating early upper air data substantially reduces uncertainties in the northern hemispheric atmospheric state back to the late 1930s. The inclusion of upper air data particularly from the tropics (Stickler et al. 2014) that have not yet been assimilated, together with measures to deal with the

strong N-S data density asymmetry can further improve results on the global scale, so that future full reanalysis efforts can be sensibly extended further backward.

Upper-air data are available back to the late 1910s in substantial number on a large (though not global) scale. Assimilating these observations will bring substantial benefit to atmospheric reanalyses. In addition, the positive results as described in this paper underline the importance of the recovery and digitization of historical data records and the impact they will have on our knowledge on the state of the atmosphere in the first part of the 20th century.

Acknowledgements. The work was funded by FP7 projects ERA-CLIM (grant agreement no. 265229), ERA-CLIM2 (grant agreement no. 607029) and the Swiss National Science Foundation project EVALUATE (130407). The contribution of M. Mayer was funded by the Austrian Wissenschaftsfonds (FWF projects P25260-N29 ,P28818-N29). 20CR data were obtained courtesy of the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the NOAA Climate Goal. The Project used resources of the National Energy Research Scientific Computing Center and of the National Center for Computational Sciences at Oak Ridge National Laboratory, which are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and Contract No. DE-AC05-00OR22725, respectively

References

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. Int. J. Climatology 6, 661-675, DOI: 10.1002/joc.3370060607

Baldwin, M. P., L. J. Gray, T. J. Dunkerton, K. Hamilton, P. H. Haynes, W. J. Randel, J. R.Holton, M. J. Alexander, I. Hirota, T. Horinouchi, D. B. A. Jones, J. S. Kinnersley, C. Marquardt, K. Sato, M. Takahashi, 2001: The Quasi-Biennial Oscillation. *Rev. Geophys.*, **39,** 179–229.

Barriopedro, D., M. Antón, J. A. García, 2010: Atmospheric blocking signatures in total ozone and ozone miniholes. *J. Climate*, **23,** 3967–3983.

Brönnimann, S., 2003: A historical upper air data set for the 1939-1944 period. *Int. J. Climatol.*, 23, 769-79.

Brönnimann, S. und J. Luterbacher, 2004: Reconstructing Northern Hemisphere upper level fields during World War II. *Climate Dynamics,* **22,** 499-510

Brönnimann, S. and G. P. Compo, 2012: Ozone highs and associated flow features in the first half of the twentieth century in different data sets. *Meteorol. Z.* **21,** 49-59.

Brönnimann S., J. Staehelin, S. F. G. Farmer, J. C. Cain, T. M. Svendby and T. Svenøe, 2003: Total ozone observations prior to the IGY. I: A history. *Q. J. Roy. Meteorol. Soc.,* **129,** 2797-2817.

Brönnimann S., J. L. Annis, C. Vogler, and P. D. Jones, 2007: Reconstructing the Quasi-Biennial Oscillation back to the early 1900s. *Geophys. Res. Lett.,* **34,** L22805.

Brönnimann, S., T. Griesser, and A. Stickler, 2012a: A gridded monthly upper-air data set from 1918 to 1957. *Climate Dynamics,* **38,** 475-493.

Brönnimann S., A. N. Grant, G. P. Compo et al. 2012b: A multi-data set comparison of the vertical structure of temperature variability and change over the Arctic during the past 100 years. *Clim. Dyn.,* **39,** 1577–1598.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X. Wang, S.D. Woodruff, and S.J. Worley, 2011: The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.,* **137,** 1-28.

Dai, A., J. Wang, P.W. Thorne, D.E. Parker, L. Haimberger, and X.L. Wang, 2011: A new approach to homogenize daily radiosonde humidity data. J. Climate, 24, 965–991.

Dobson, G. M. B. and D. N. Harrison, 1926: Observations of the amount of ozone in the Earth's atmosphere and its relation to other geophysical conditions. *Proc. R. Soc. A,* **110,** 660–693.Eskridge, R. E., Alduchov, O. A., Irina V. Chernykh, I. V., Zhai Panmao, Z., Arthur C. Polansky, A. C. and Doty, S. R., 1995: A Comprehensive Aerological Reference Data Set (CARDS): Rough and Systematic Errors. Bull. Amer. Meteorol. Soc. 76, 1759-1775.

Feuchter, D., M. Mosimann, M. Tscherrig (2014) 1944 Typhoon Cobra in 20CR, ERA-20C and ERA-PreSAT reanalysis data. Student Paper, University of Bern, 13 pp.

Grant, A., S. Brönnimann and T. Ewen, 2009: A New Look at Radiosonde Data prior to 1958. *J. Climate.,* **22,** 3232-3247

Griesser, T., S. Brönnimann, A. Grant, T. Ewen, A. Stickler, and J. Comeaux, 2010: Reconstruction of global monthly upper-level temperature and geopotential height fields back to 1880. *J. Climate,* **23,** 5590-5609

Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. J. Climate 20, 1377-1403.

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, and D. Dee, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble. *Quarterly Journal of the Royal Meteorological Society.*

Kistler, R. and co-authors, 2001: The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *B. Amer. Meteorol. Soc.*, **82,** 247-267

Kobayashi, S. et al., 2015: The JRA-55 reanalysis: General specifications and basic characteristics. J. Meteorol. Soc. Jpn. 93, 5–48.

Mayer, M., and L. Haimberger, 2012: Poleward Atmospheric Energy Transports and Their Variability as Evaluated from ECMWF Reanalysis Data. *J. Climate*, **25**, 734–752. doi: http://dx.doi.org/10.1175/JCLI-D-11-00202.1

Orsolini, Y. J., D. B. Stephenson, and F. J. Doblas-Reyes, 1998: Storm tracks signature in total ozone during Northern Hemisphere winter. *Geophys. Res. Lett.*, **25,** 2413–2416.

Poli, P. et al., 2015: ERA-20C Deterministic. ERA Report Series 20, ECMWF, Reading

Poli, P., H. Hersbach, D. P. Dee, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. H. Tan, C. Peubey, J-N. Thepaut, Y. Tremolet, E. V. Holm, M. Bonavita, L. Isaksen and M. Fisher : 2016: ERA-20C : An atmospheric reanalysis of the 20th century. *J. Clim.*, **29,** 4085-4097.

Ramella-Pralungo, L.R., Haimberger, L., Stickler, A., Brönnimann, S., 2014: A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 - Part 1: Merging and interpolation to 00:00 and 12:00 GMT. Earth System Science Data, 6, 185-200.

Ramella Pralungo, L., and L. Haimberger, 2015: New estimates of tropical mean temperature trend profiles from zonal mean historical radiosonde and pilot balloon wind shear observations, J. Geophys. Res. Atmos., 120, 3700–3713.

Rayner, N., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.

Saha, S. and co-authors, 2010: The NCEP Climate Forecast System Reanalysis. *B. Amer. Meteorol. Soc.* **91,** 1015-1057.

Simmons, A., Hersbach, H., Poli, P., Bidlot J. et al. 2015: Revisiting the weather around D-Day using a modern analysis and forecasting system and recently and forecasting system and recently digitized historical digitized historical observations.http://www.rmets.org/sites/rmets.org/files/presentations/07022015-hersbach.pdf

Staehelin J., A. Renaud, J. Bader, R. McPeters, P. Viatte, B. Hoegger, V. Buignon, M. Giroud, and H. Schill, 1998: Total ozone series at Arosa (Switzerland): Homogenisation and data comparison. *J. Geophys. Res.*, **103,** 5827–5841.

Stickler A., A. N. Grant, T. Ewen et al. 2010: The comprehensive historical upper-air network. *Bull. Am. Meteorol. Soc.*, **91,** 741–751.

Stickler, A., Brönnimann, S., Jourdain, S., Roucaute, E., Sterin, A., Nikolaev, D., Valente, M. A., Wartenburger, R., Hersbach, H., Ramella-Pralungo, L., and Dee, D. (2013) Description of the ERA-CLIM historical upper-air data, *Earth Syst. Sci. Data*, **6,** 29-48, doi:10.5194/essd-6-29-2014.

Stickler, A., S. Brönnimann, M. A. Valente, J. Bethke, A. Sterin, S. Jourdain, E. Roucaute, M. V. Vasquez, D. A. Reyes, R. Allan, and D. Dee (2014) ERA-CLIM: Historical Surface and Upper-Air Data for Future Reanalyses. *B. Amer. Meteorol. Soc.*, **95,** 1419–1430.

Stickler, A., S. Storz, C. Jörg, R. Wartenburger, H. Hersbach, G. Compo, P. Poli, D. Dee, and S. Brönnimann (2015) Upper‐air observations from the German Atlantic Expedition (1925-27) and comparison with the Twentieth Century and ERA‐20C reanalyses. Meteorol. Z., PrePub Article, doi:10.1127/metz/2015/0683.

Uppala, S. M., and co-authors, 2005: The ERA-40 re-analysis. *Q. J. Roy. Meteorol. Soc.*, **131,** 2961-3012.

Vaughan, G. and J. D. Price, 1991: On the relation between total ozone and meteorology. - *Q. J. Roy. Meteorol. Soc.*, **117,** 1281-1298.

Vogler, C., S. Brönnimann, J. Staehelin and R. E. M. Griffin (2007) The Dobson total ozone series of Oxford: Re-evaluation and applications. *J. Geophys. Res.* **112**, D20116, doi:10.1029/2007JD008894.

Wartenburger, R., S. Brönnimann, and A. Stickler, 2013: Observation errors in early historical upper-air observations, J. Geophys. Res. 118, 12012-12028.

Wegmann, M, S. Brönnimann and G. P. Compo (2016) Tropospheric circulation during the early twentieth century Arctic warming. *Clim. Dyn.*, published online, doi:10.1007/s00382-016-3212-6.

Table 1. List of the historical upper-air data sets input to ERA-PreSAT. The observables Temperature (T), wind speed (Ws), wind direction (Wd), specific humidity (Q), relative humidity (R), dew-point depression ($T-T_d$) and vertical coordinates pressure (P) and height (Z) are not available for all stations nor the entire period, especially for humidity and dew-point depression.

Dataset	**Location**	**Available Period**	**Available Observables**
NCAR UADB-2	http://rda.ucar.edu/datasets/ds370.1	Feb 1919 – Aug 2012	Z, P, T, Ws, Wd, RH
CHUAN v1.7 'Raw'	http://rda.ucar.edu/datasets/ds370.1	Jan 1904 – Mar 2007	Z, P, T, Ws, Wd, Q
CHUAN v1.7 'Corrected'	http://rda.ucar.edu/datasets/ds370.1	Jan 1904 – Mar 2007	Z, P, T, Ws, Wd
ERA-CLIM v0.9	University Bern	Oct 1899 – Dec 1972	Z, P, T, Ws, Wd, Q, RH, $T-T_d$

Table 2. Historical total ozone stations used in this study (see Brönnimann et al. 2003). n denotes the number of days with observations within the period 1924-1963.

Station	Period	lon (° E)	lat (° N)	n
Aarhus	1952-1963	10.6	56.3	3462
Aldergrove	1952-1957	-6.2	54.7	1415
Arosa	1939-1963	9.7	46.8	6438
Camborne	1952-1953	-5.3	50.2	3283
College	1952-1957	-147.5	64.7	394
Dombas	1940-1946	9.1	62.1	1410
Edmonton	1950-1952	-113.5	53.6	557
Flagstaff	1954-1957	-111.7	35.2	463
Gulmarg	1955-1956	74.4	34.1	298
Hemsby	1952-1955	1.7	52.7	855
Lerwick	1939-1963	-1.2	60.1	3290
Magny	1955-1959	2.1	48.7	1054
Mount Abu	1951-1960	72.7	24.6	2278
New Delhi	1955-1957	77.2	28.6	920
New York	1941-1944	-73.9	40.9	899
Oxford	1939-1963	-1.2	51.8	6020
Rome	1954-1963	12.2	42.1	3152
Spitsbergen	1950-1962	15	78	1676
Srinagar	1956-1957	74.8	34.1	182
Tateno	1955-1957	140.1	36.1	409
Uppsala	1952-1963	17.6	59.9	2472

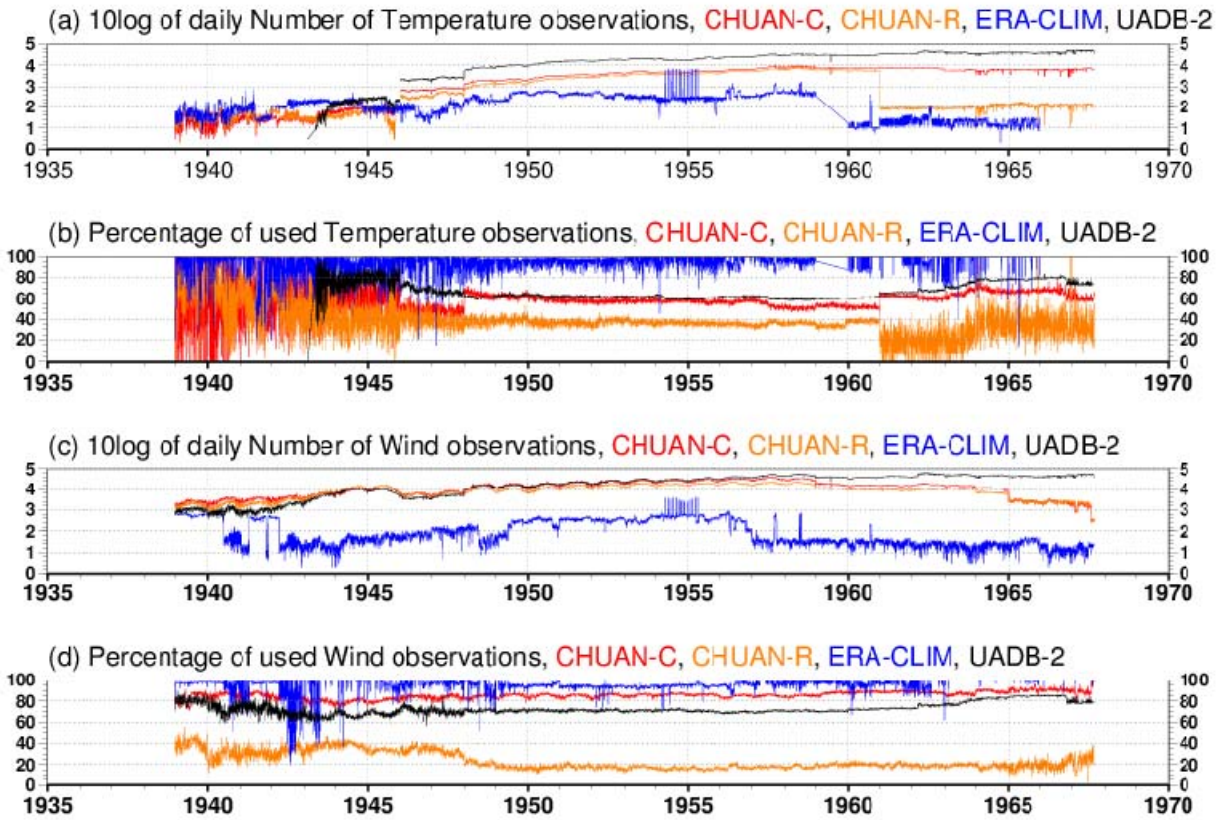
Table 3. Month-to-month Pearson correlations (x100) with upper-air data at four locations in North America, spanning latitudes between 17.5° N and 46.5 ° N, of three reconstructions (top three rows) and three reanalyses (bottom three rows) for the 1939-1944 period. Correlations are given for temperature (T) and geopotential height (Z) at the three levels 700, 500, and 300 hPa. Monthly anomalies were expressed as deviations from the mean annual cycle over the 1939-1944 period,

	Sault St. Marie, 46.5°N, n=66		Washington DC, 38.9°N, n=66		Miami, 25.8°N, n=62		Islas Santanilla, 17.5°N, n=48	
	Z300/500/700	T300/500/700	Z300/500/700	T300/500/700	Z300/500/700	T300/500/700	Z300/500/700	T300/500/700
BL2004	83/89/86	83/89/88	96/95/95	91/93/94	92/93/94	67/86/91	77/71/40	74/63/22
REC1	82/86/79	77/88/88	93/92/91	82/89/89	83/88/89	53/82/74	26/26/33	43/35/47
REC2	89/93/93	77/93/93	96/96/96	80/91/93	91/91/92	84/87/86	76/64/57	80/72/72
ERA-PreSAT	84/91/93	71/80/89	87/91/93	62/79/90	60/80/91	37/55/61	-06/-07/36	08/06/39
ERA-20C	73/82/87	53/71/79	87/88/90	74/78/84	73/81/80	33/63/61	29/24/48	28/24/45
20CRv2	80/86/89	70/83/85	84/90/92	57/75/85	68/80/83	46/60/54	-01/09/39	-13/05/27

Table 4. Pearson correlations of Dec.-Feb. mean temperature at 700 hPa north of 60° N from 1940 to 1961 in different data sets (see Wegmann et al. 2016)

	20CRv2	ERA-20C	ERA-PreSAT	REC1
20CRv2	1	0.77	0.96	0.88
ERA-20C	0.77	1	0.80	0.68
ERA-PreSAT	0.96	0.80	1	0.87
REC1	0.88	0.68	0.87	1

Figures

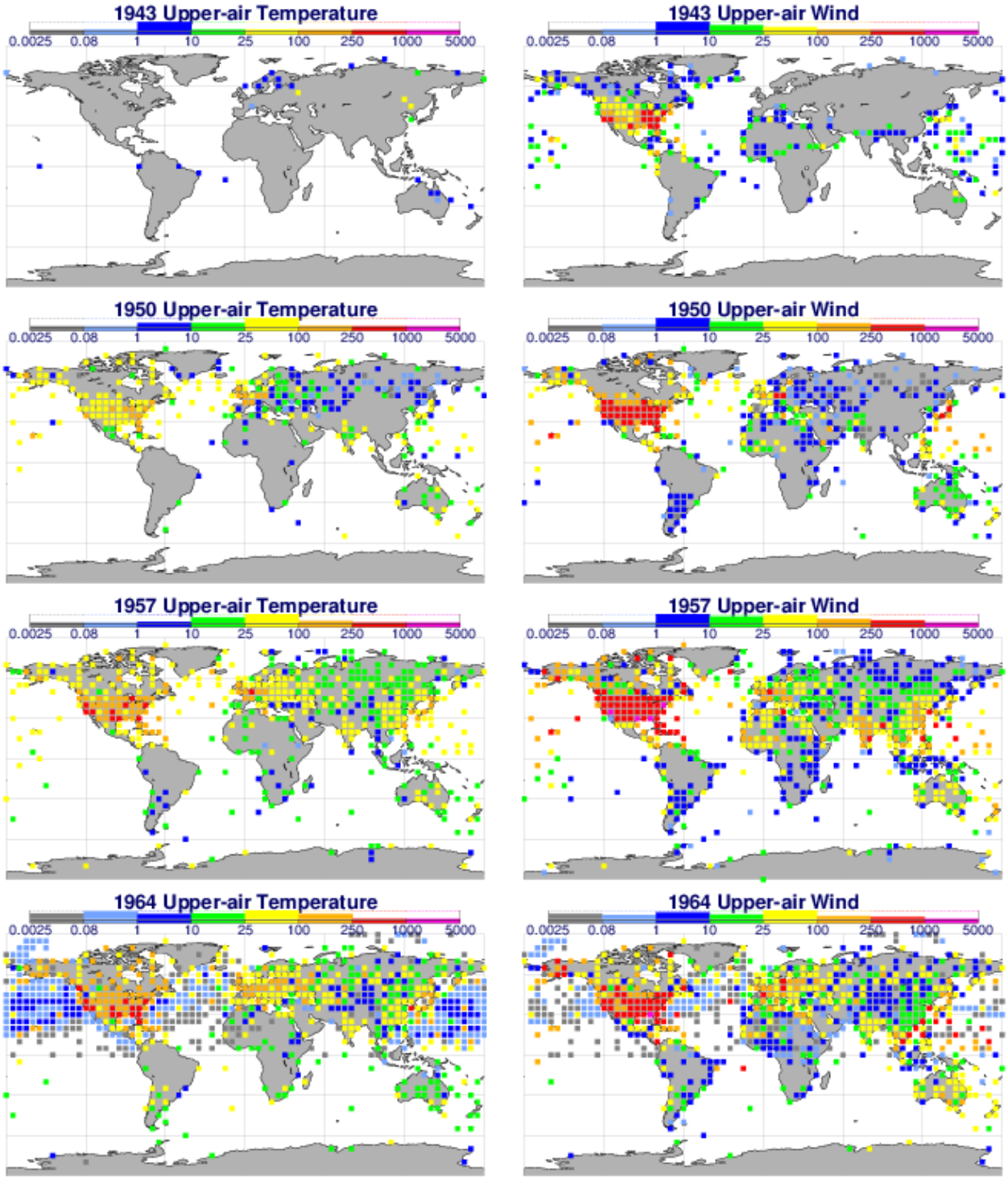


738

739 **Fig.1:** Time series of availability and usage of upper-air data in ERA-PreSAT for the
740 CHUAN v1.7 'corrected' (red)), CHUAN v1.7 'raw' (orange), ERA-CLIM v0.9
741 (blue) and UADB-2 (black) data sets.

742

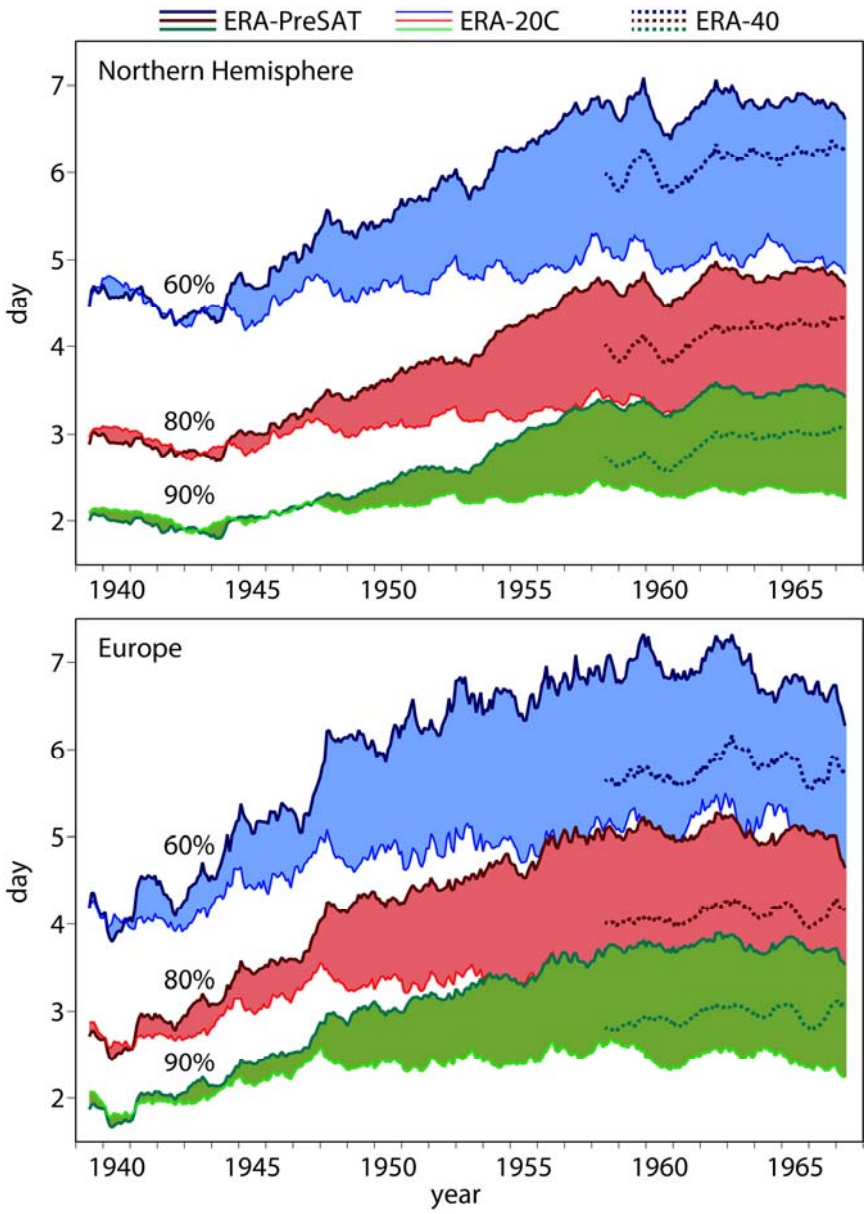
743



744

745 **Fig. 2:** Geographical distribution and number of observations actively assimilated in
746 ERA-PreSAT per day in 5x5 degree grid boxes, averaged over (from top to bottom)
747 1943, 1950, 1957 and 1964, for upper-air temperature (left) and wind (right). The
748 lowest two contour bounds of 0.0025 and 0.08 represent one observation in one entire
749 year and one observation per month, respectively.

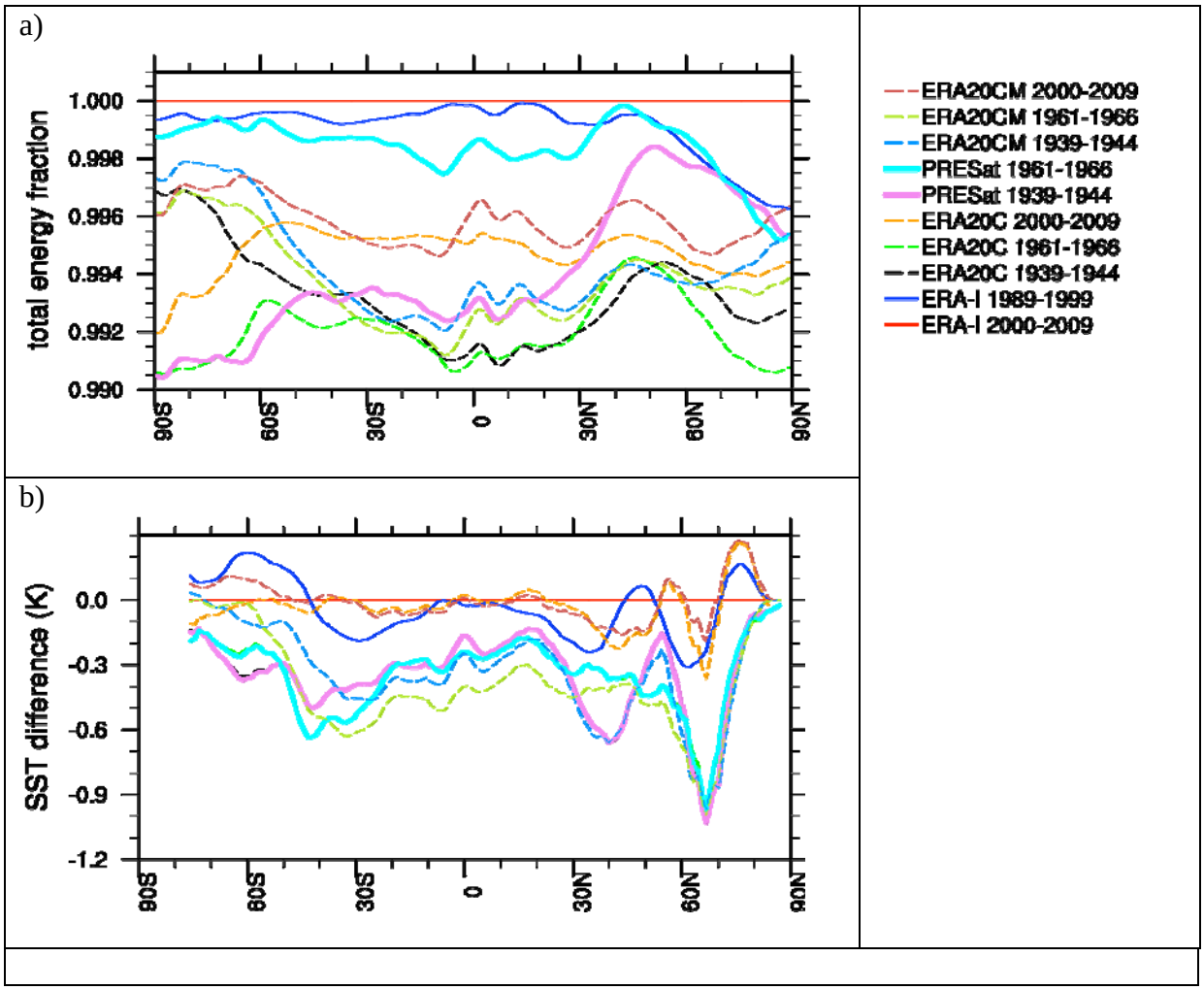
750



751

752 **Fig. 3**: Twelve-month running mean of the number of forecast days for which the
753 northern hemispheric (top panel) and European (lower panel) anomaly correlation
754 coefficient (geopotential at 500 hPa height) with respect to the own verifying analysis
755 reaches 90% (green), 80% (red) and 60% (blue) for ERA-PreSAT (dark solid curves),
756 ERA-20C (light solid curves) and ERA-40 (dotted curves). All forecasts started from
757 3 hours into the assimilation window, which means forecasts for ERA-20C and ERA-
758 PreSAT (24-hour assimilation window) have benefitted from observations that were
759 21 hours into the future while only 3 hours in ERA-40 (6-hour window).

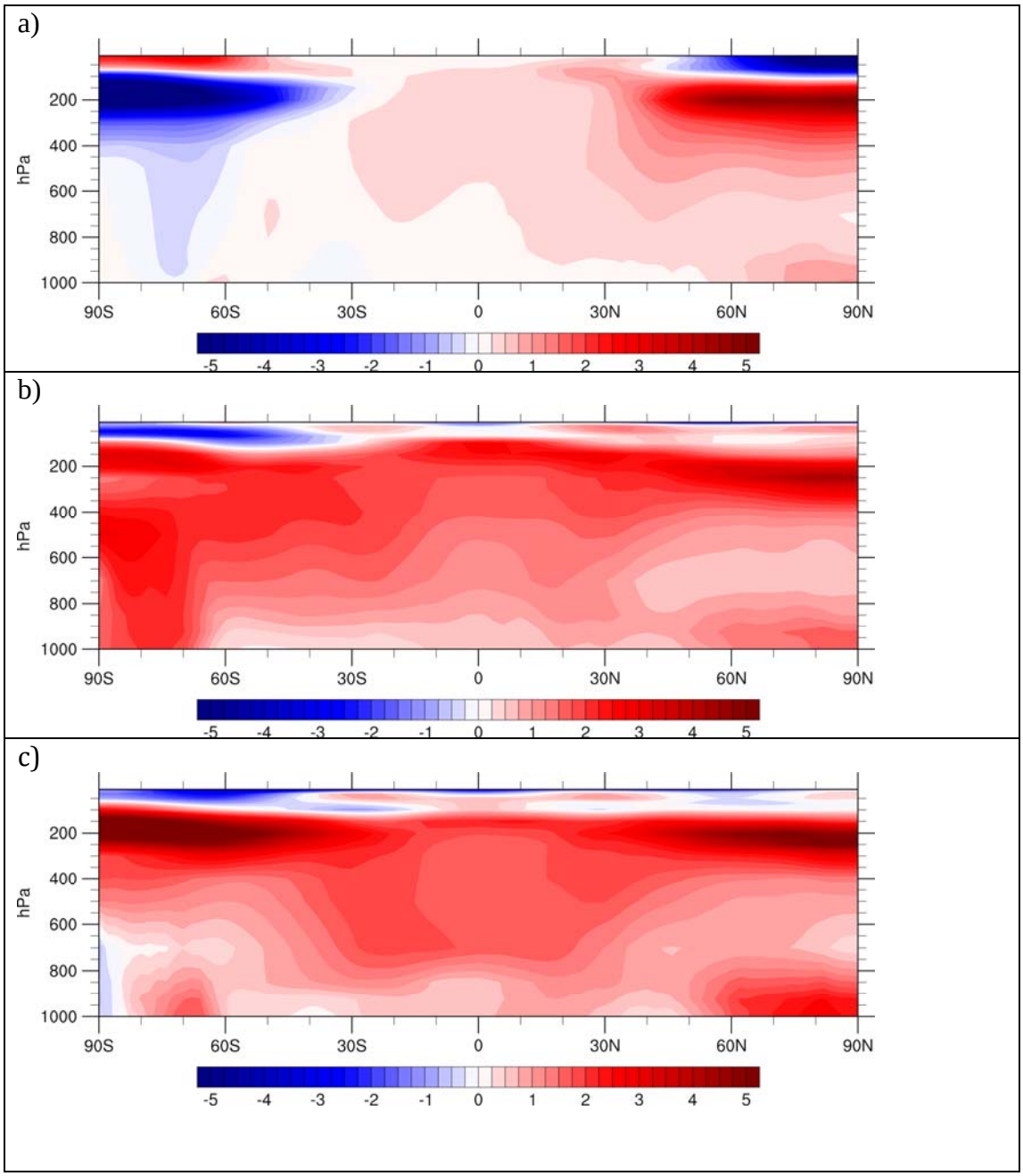
760



761

762 **Fig. 4** a) Fraction of zonal mean specific total energy from
763 ERA20CM/ERA20C/ERA-PreSAT/ERA-I (ERA-Interim 1989-1999) and ERA-I for
764 the reference period (2000-2009); b) Difference of zonal mean SSTs from ERA-I
765 (2000-2009) and ERA20CM/ERA20C/ERA-PreSAT/ERA-I (1989-1999); The
766 different averaging periods are given in the legend.

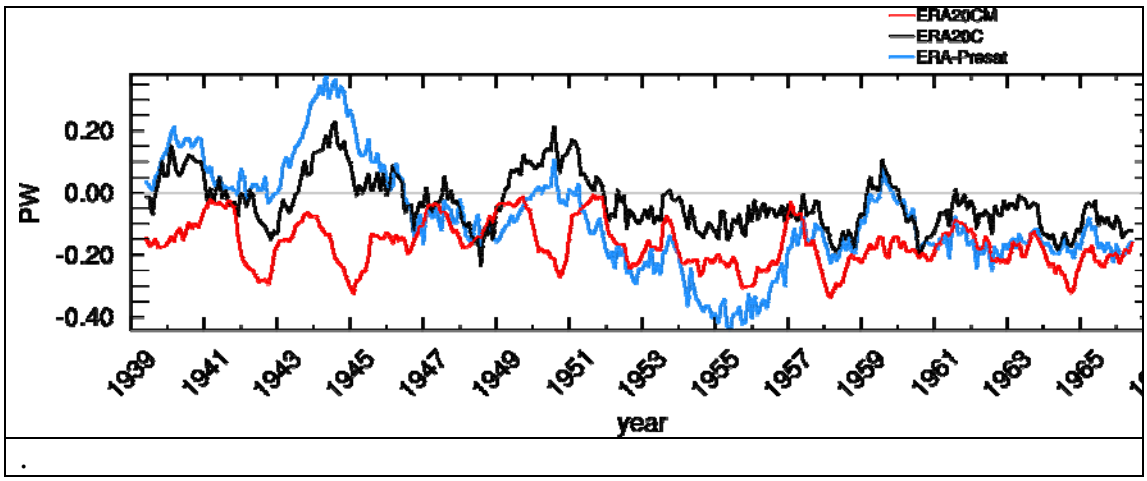
767



768

769 **Fig. 5** Difference of ERA-PreSAT and ERA-20C zonal mean temperatures for a)
770 1939-1944 and b) 1961-1966, and c) differences ERA-Interim-ERA-20C 1980-1989,
771 Units are K.

772
773



.

774

775 **Fig 6.** 12-month running mean total energy transport across the equator (positive
776 northward) from ERA20CM, ERA20C, and ERA-Presat. The ERA-I reference value
777 is -0.25PW.

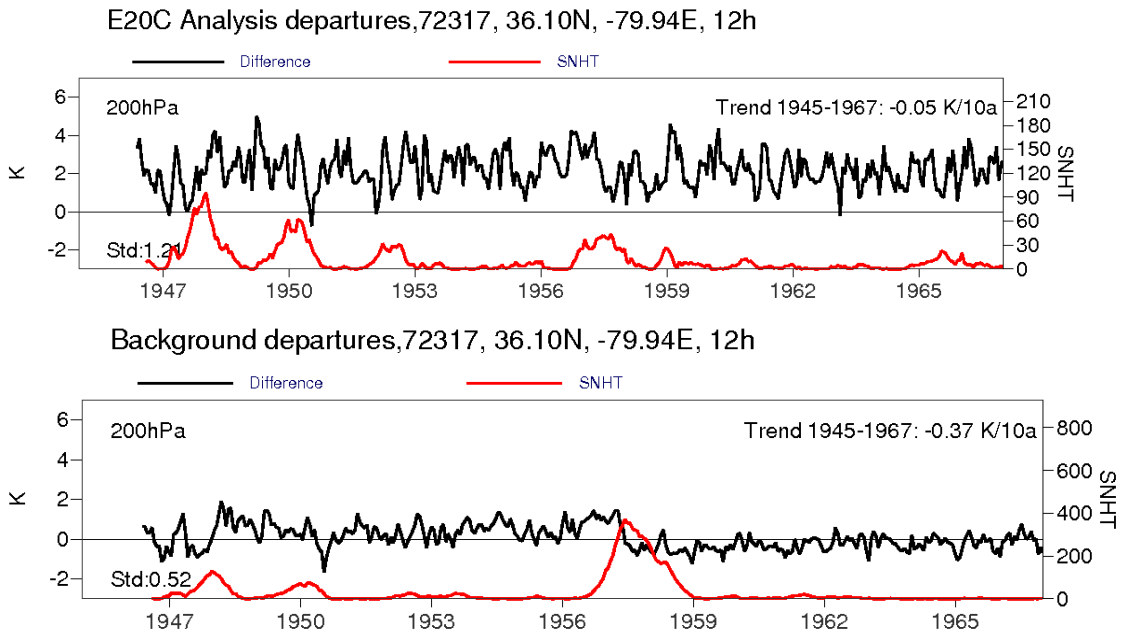
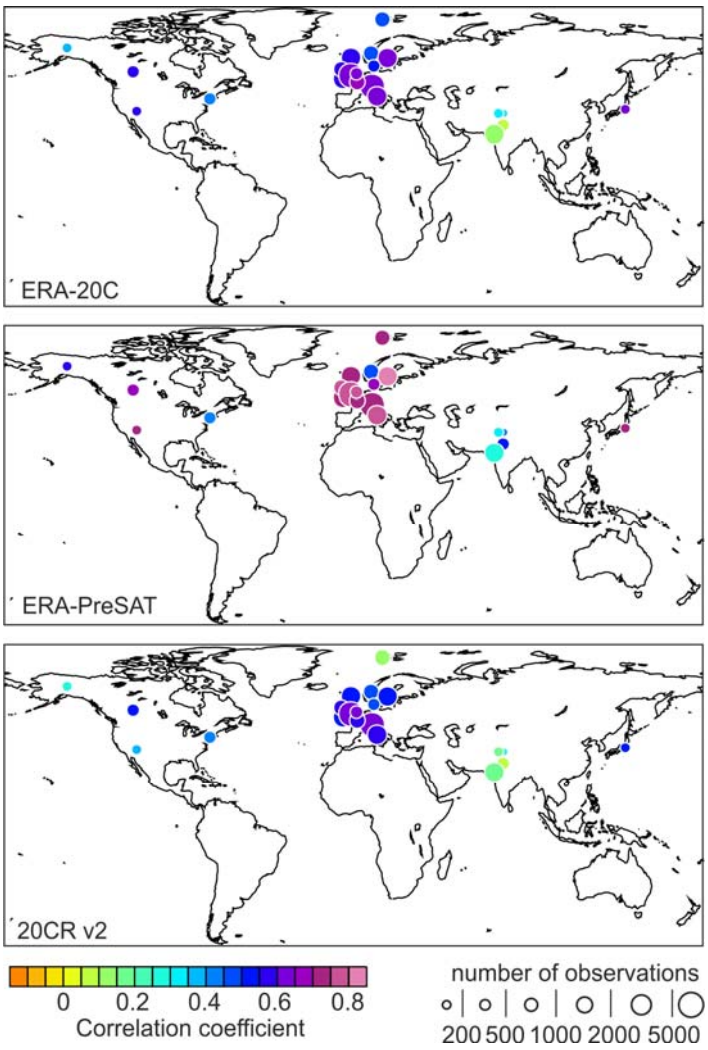


Fig. 7*:* Time series of obs-ERA-20C analysis departures (upper panel, standard deviation 1.21K) and obs-ERA-preSAT background departures (lower panel, standard deviation 0.52K) for station Greensboro in the eastern US. Red curve (right axis) is Standard Normal Homogeneity Test statistic as described in Haimberger (2007). Sharp maxima with values above 50 indicate likely breakpoints. Note good temporal correspondence of maxima in both panels, detection efficiency is better in lower panel due to smaller noise level.



790

791

792 **Fig. 8**: Correlations between total ozone anomalies from reanalysis data sets and
793 observations, 1939-1963. The size of the circle indicates the number of observations
794 for the stations listed in Table 1 (see Brönnimann and Compo, 2012 for correlations
795 with 300 hPa geopotential height).

796

797

798

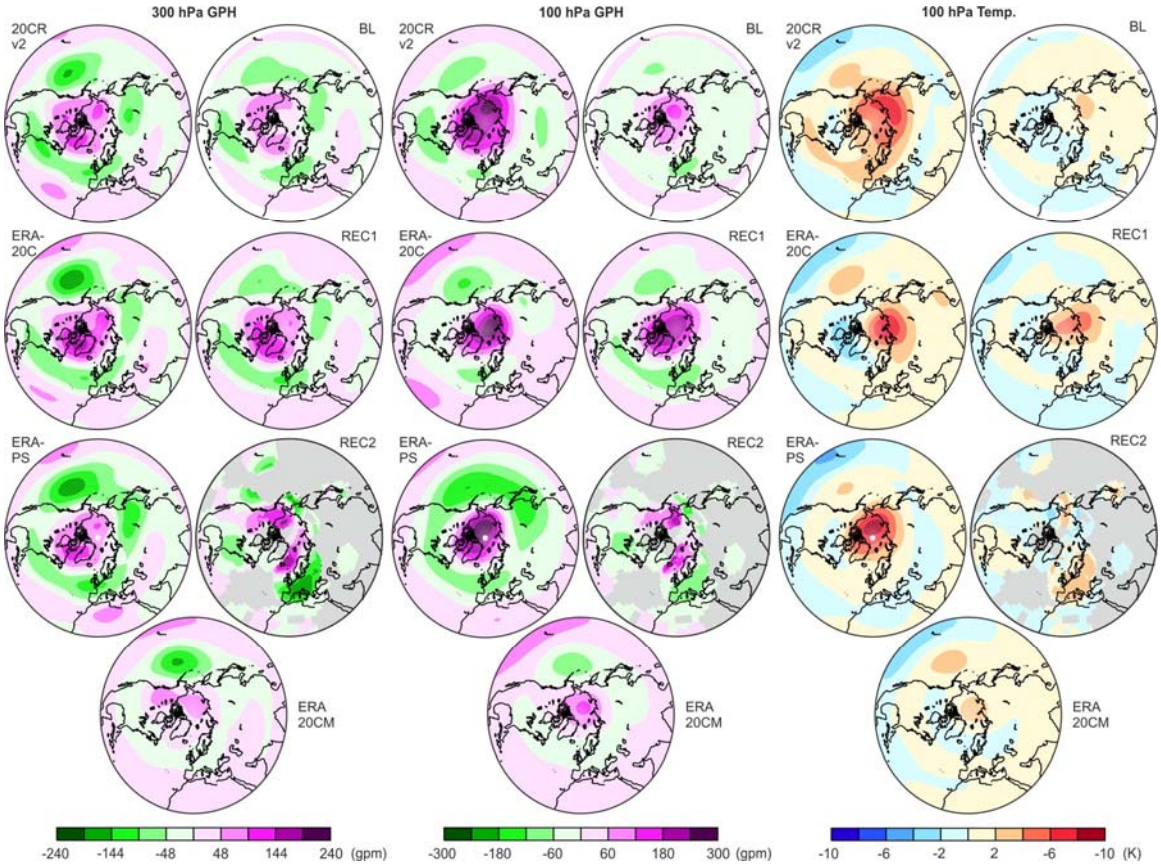


Fig. 9: Difference in (left) 300 hPa GPH, (middle) 100 hPa GPH and (right) 100 hPa temperature over the northern extratropics between the January to April period of 1940-1942 and that of the neighbouring years (1939, 1943 1944) from different data sets.

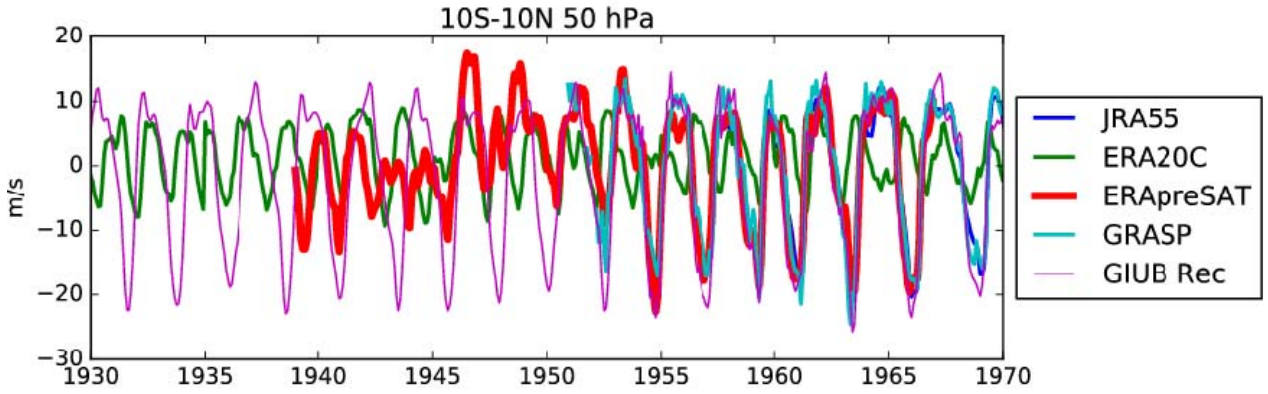


Figure 10: Time series of 50hPa u-wind component averaged zonally and over the tropical belt for different reanalyses, one wind observation data set (GRASP) and the reconstruction of Brönnimann et al. 2007 (GIUB Rec). Note that ERA-PreSAT is the only reanalysis that captures QBO from late 1940s up to 1957.

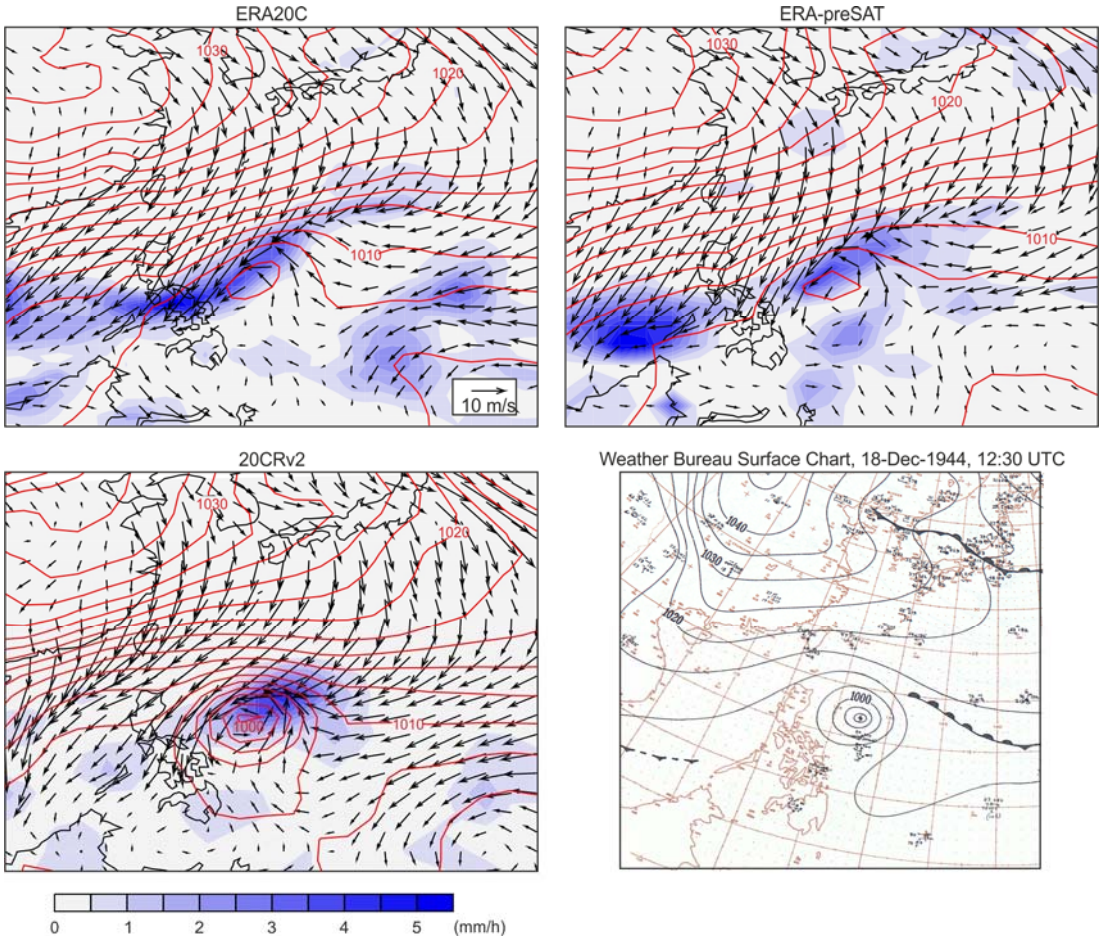


Fig. 11: Fields of sea-level pressure, 10 m wind, and precipitation on 18 Dec. 1944, 6 UTC from four different reanalyses. The bottom right panel shows the surface analysis from the Weather Bureau.